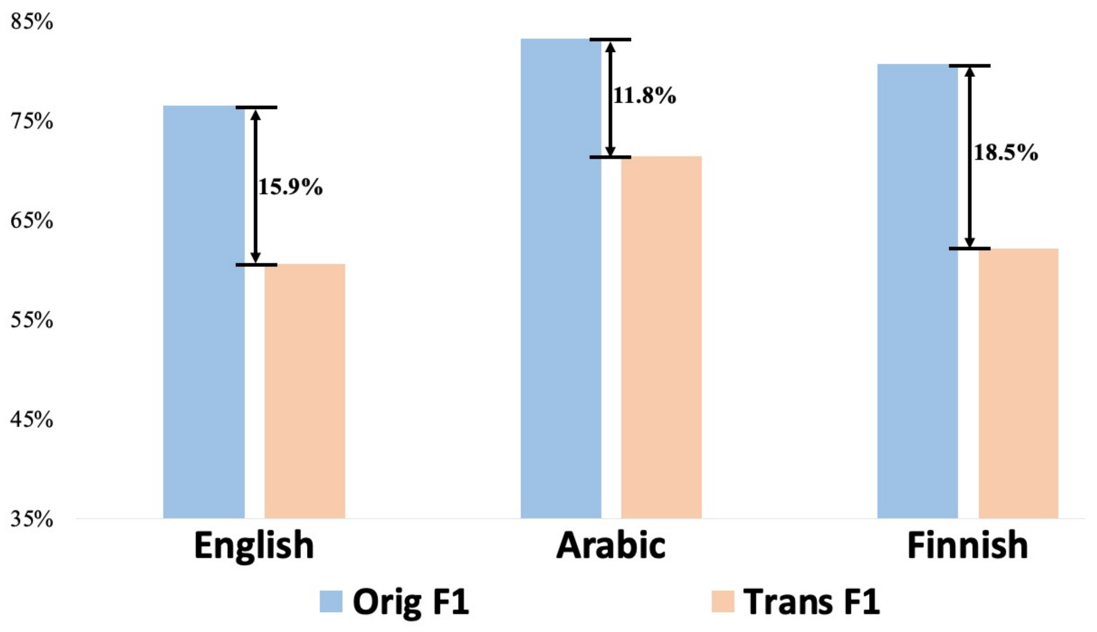
SMU SINGAPORE MANAGEMENT UNIVERSITY
School of Computing and Information Systems

Translate-Train Embracing Translationese Artifacts

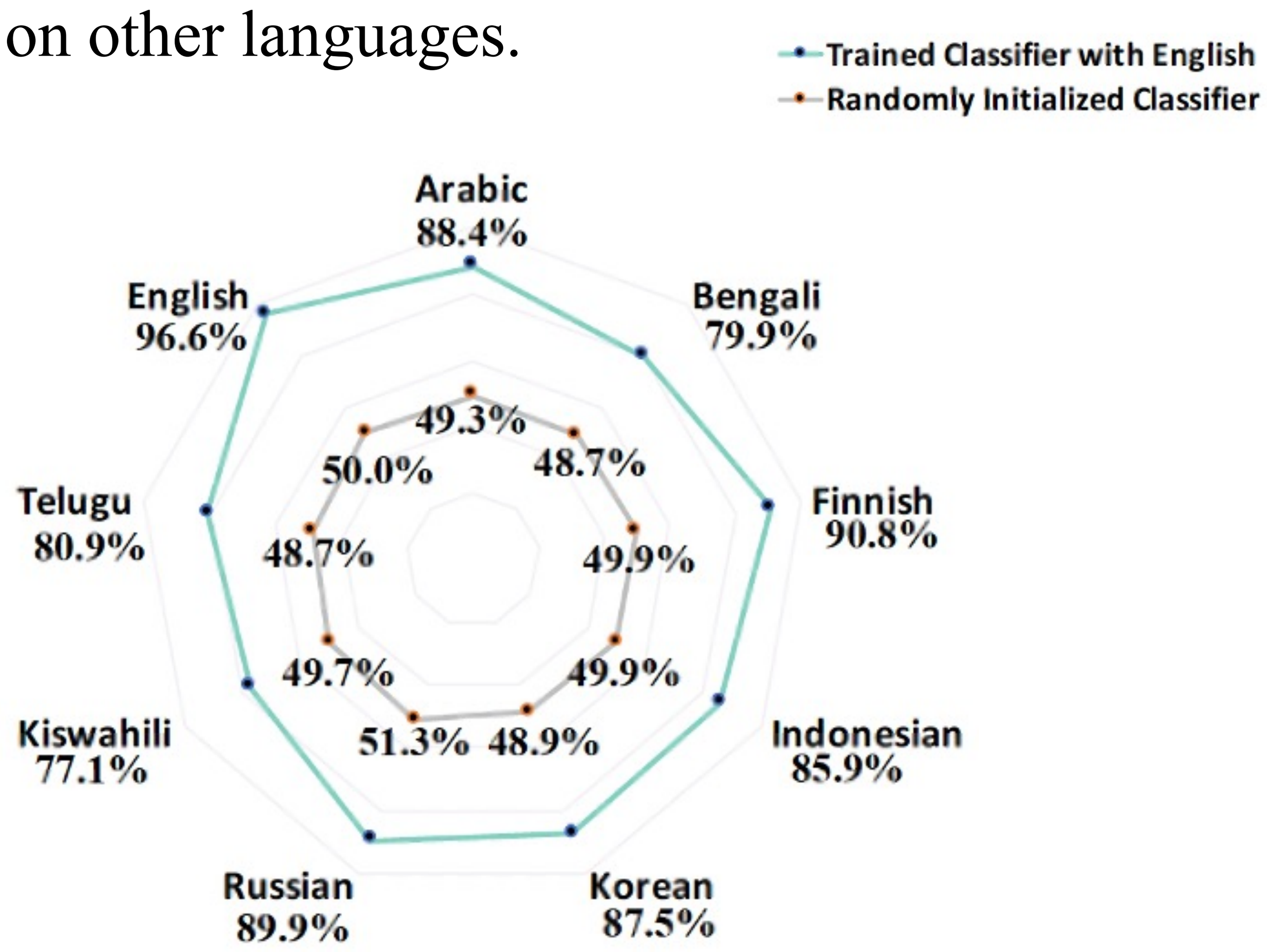
Sicheng Yu, Qianru Sun, Hao Zhang, Jing Jiang

Motivation

- Although translate-train mitigate the gap of unseen target languages, translated text, i.e., translationese, brings another gap into the model.



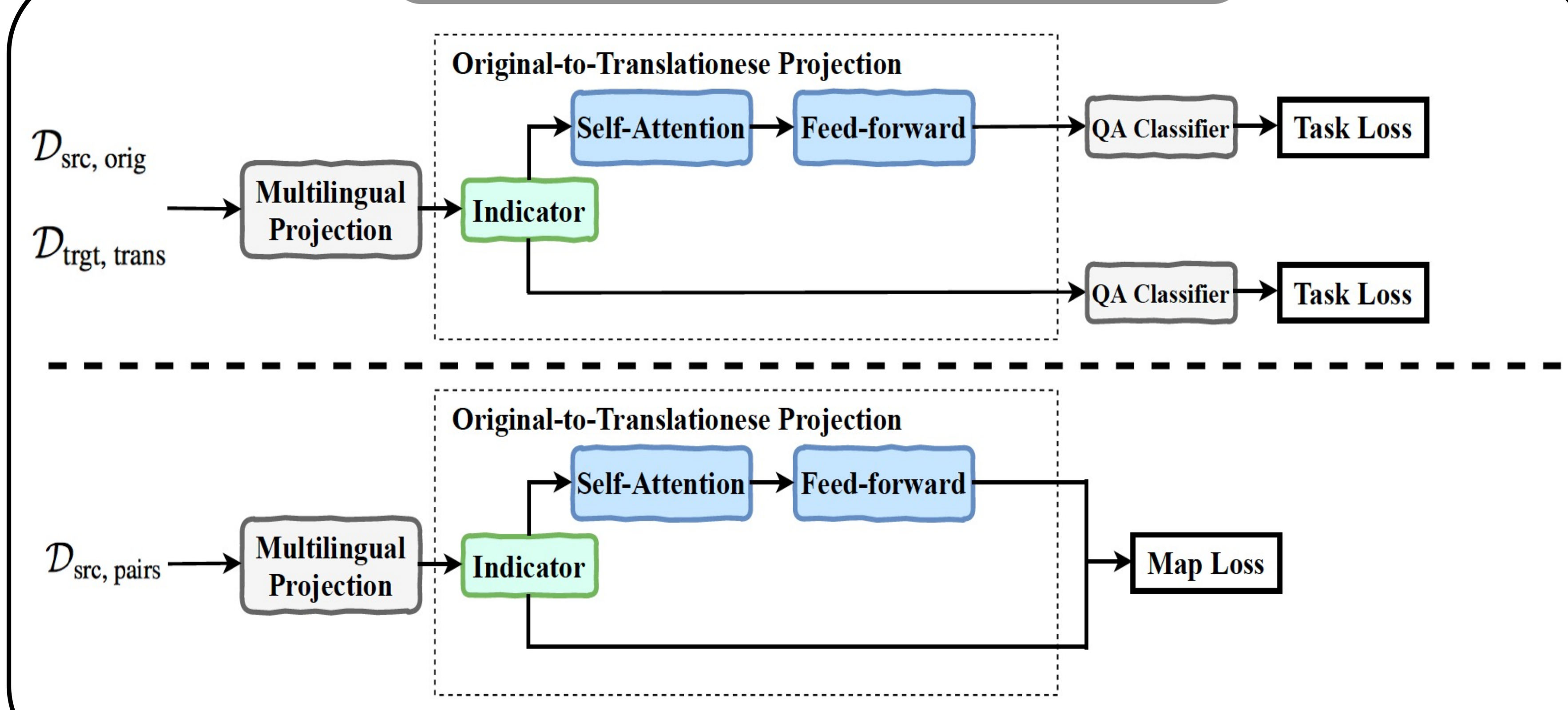
- We make a hypothesis that the originals-translationese gap in English can be generalized to other languages.
- To demonstrate it, we train a classifier to distinguish two types of data using English data and directly apply on other languages.



Our Approach

- We build on top of XLM-R and propose a originals-to-translationese projection to map the originals and translationese into the same space.
- Originals-to-translationese projection contains a self-attention layer and a feed-forward layer for originals data and no operation for translationese data.
- Translationese of English is generated to further enhance the originals-to-translationese projection.

Solution Architecture



Experiment Results

Method	D	ar	bn	fi	id	ko	ru	sw	te	avg
STT	✘	40.4/67.6	47.8/64.0	53.2/70.5	61.9/77.4	10.9/31.9	42.1/67.0	48.1/66.1	43.6/70.1	43.5/64.3
FILTER	✘	50.8/72.8	56.6/70.5	57.2/73.3	59.8/76.8	12.3/33.1	46.6/68.9	65.7/77.4	50.4/69.9	49.9/67.8
STT*	✘	58.0/76.6	54.6/70.2	59.0/74.8	64.7/80.2	48.0/61.6	49.5/ 71.2	58.7/74.6	57.0/76.2	56.2/73.2
TAG*	✔	56.9/76.4	55.5/70.0	59.4/75.2	64.4/79.6	48.6/61.7	49.1/70.4	60.7/76.0	57.8/76.4	56.5/73.2
TST*	✔	58.4/75.5	60.2/72.2	58.3/74.4	65.5/78.9	49.3/62.6	49.0/69.7	63.5/76.7	56.2/76.1	57.6/73.3
GRL*	✔	57.6/75.6	58.4/72.6	59.7/74.8	65.3/79.9	49.6/62.2	49.1/70.4	62.9/76.9	58.2/77.0	57.6/73.7
TEA*	✔	56.5/76.1	60.2/74.9	60.9/76.5	63.6/79.3	48.6/61.4	51.5/72.0	66.7/78.9	60.7/78.7	**58.6/74.7**

Table 1: Main results (Exact Match / F1 scores) on TyDiQA. All methods are with XLM-R as backbone. The "D" column indicates whether the design of this method considers translationese artifacts. The columns "ar" to "te" represent different target languages. The "avg" column denotes the average performance across the 8 target languages. * indicates our implementation.

- Methods considering the gap between translationese and originals perform better.
- Our method surpasses strong baselines.

Settings	EM	F1
STT	56.2	73.2
(1) STT+$\mathcal{X}_{\text{src, trans}}$	56.6	73.2
(2) STT+params	56.3	73.5
(3) TOP	57.9	74.1
(4) MLP in OTP	56.7	73.3
(5) MSE loss	58.0	73.9
Full method	**58.6**	**74.7**

Table 2: Ablation study on TyDiQA. We report the average EM and F1 performance on the 8 target languages.

- The improvement of our method is not caused by additional parameters or data.
- TOP still mitigates the artifacts, but OTP obtaining better performance.
- Our loss function and architecture are more effective.

Settings	Language Family	EM	F1
Scottish (gd)	Indo-European	58.8	74.0
Korean (ko)	Koreanic	57.8	74.0
Chinese (zh)	Sino-Tibetan	57.6	73.8
German (de)	Indo-European	58.6	74.7

Table 3: Experiment results of utilizing different language as pivot language for generating $\mathcal{X}_{\text{src, trans}}$.

- Pivot languages from Indo-European family are superior to that from other language families.

Conclusions

- We expose the drawback caused by translationese in translate-train and demonstrate that the pattern of translationese is transferrable.
- We propose a simple mapping method learned on English to mitigate the translationese artifacts.

Reference

Sicheng Yu, Qianru Sun, Hao Zhang, Jing Jiang. Translate-Train Embracing Translationese Artifacts. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.